AI Course

Capstone Project   
Action Plan

For students (instructor’s review required)

ⓒ2023 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of this document.

This document is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this document other than the curriculum of Samsung Innovation Campus, you must receive written consent from copyright holder.

|  |  |
| --- | --- |
| Course | AI Course |
| Team Name | Outliers (Group 4) |
| Team Leader/  Members | Ranim Alfaraj /  Ahmad Alghali, Madhumitha Ichapuram, Abdelwahid Eltayeb, Taha Nasir, Mustafa Eltayeb |
| Project Title | SmartMeds Guide: A Drug Recommendation System |
| Goal |  |
| Developing a system to prescribe drugs based on patients' symptoms involves creating an intelligent and reliable platform that leverages advanced technologies, data analysis, and medical expertise to enhance the accuracy and efficiency of healthcare. | |
| Abstract |  |
| Everyday we are seeing new advancements in technology contributing to the improvement of society in many fields, one of those fields being medicine. There are instances when patients are unable to schedule medical appointments due to several reasons, and therefore require alternatives for the traditional methods of medical assistance. Machine learning has made life much easier and is nowadays being implemented in many aspects of our daily lives. Our project aims to create a recommender system that prescribes drugs based on the symptoms a patient might be feeling. The dataset used in this project provides patient reviews on specific drugs, related conditions the patient had been suffering from, and a 10-star patient rating reflecting overall patient satisfaction. | |
| Method |  |
| Multiple machine learning techniques such as SVM, KNN and rule-based systems can be combined. Support Vector Machine (SVM) can be used for drug recommendation due to its effectiveness in multi-class classification and ability to handle high-dimensional data. K-Nearest Neighbors (KNN) can serve as a secondary method for refining recommendations and addressing limitations of SVMs in specific scenarios. Rule-Based System will integrate medical knowledge and clinical guidelines to ensure recommendations align with established best practices. The data is already divided into training and testing data. The SVM and KNN models will be trained on the training data, optimizing hyperparameters based on the validation set. The performance of the models will be evaluated using the testing data. **1. Data Preparation:**  * Cleaning: Remove HTML tags, convert to lowercase, remove stop words, and apply lemmatization (consistent with article's approach). * Vectorization: Use TF-IDF to represent text reviews as numerical vectors. * Splitting: Divide data into training and testing sets (e.g., 80/20 split).  **2. SVM (Support Vector Machine):**  * Model Selection: Choose a suitable kernel (e.g., linear, RBF) for representing relationships between features and medical conditions. * Training: Train the SVM model on the training set to learn decision boundaries for predicting medical conditions. * Prediction: Use the trained model to predict medical conditions for new reviews. * Considerations: SVMs can handle high-dimensional data and non-linear relationships well, but might require careful hyperparameter tuning and can be computationally expensive.   **3. KNN (K-Nearest Neighbors):**   * Distance Metric: Select an appropriate distance metric (e.g., Euclidean, cosine similarity) to measure similarity between reviews. * Parameter Tuning: Experiment with different values of k (number of neighbors) to find the optimal setting. * Prediction: For a new review, identify its k nearest neighbors in the training set and assign the most common medical condition among those neighbors. * Considerations: KNN is simple to implement and interpret, but might be less efficient for large datasets and sensitive to feature scaling.   **4. Rule-Based System:**   * Rule Creation: Collaborate with medical experts to define rules based on symptoms, drug interactions, patient demographics, and other relevant factors. * Inference: Use these rules to infer medical conditions and recommend drugs based on new reviews. * Considerations: Rule-based systems offer interpretability and can incorporate domain knowledge, but require manual rule creation and maintenance, and might not generalize well to unseen cases.  **5. Hybrid Approach:** Combine SVM or KNN with Rule-Based System: Use ML models for initial predictions and then refine recommendations using rule-based checks for consistency, safety, and adherence to best practices. **6. Evaluation:**Accuracy: Measure how well each model predicts correct medical conditions.Precision and Recall: Assess how well each model identifies relevant drugs for each condition.User Feedback: Gather patient feedback on effectiveness and clarity of recommendations.Iterate: Based on evaluation results, refine models, rules, or data preprocessing steps. | |

|  |  |
| --- | --- |
| Data |  |
| The dataset of the team’s choice is from the UC Irvine Machine Learning Repository. The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction. The data is split into a train (75%) a test (25%) partition (see publication) and stored in two .tsv (tab-separated-values) files, respectively, and it has no missing values. It has 215063 instances and 7 features. | |
| Expected  Outcome |  |
| This project equips patients to manage minor ailments independently, mitigating reliance on costly medical consultations and reclaiming valuable time. By enabling the self-resolution of non-critical health concerns, independent of waiting rooms and appointments, patients achieve not only financial benefits but also the restoration of precious personal time. Furthermore, this project provides medical professionals with the opportunity to refine treatment recommendations through the utilization of patient feedback, thereby attaining a comprehensive understanding of drug efficacy across diverse demographic groups. Notably, this initiative extends beyond immediate personal advantages, encouraging patients with self-management skills and fostering a preventive healthcare approach, potentially mitigating the future necessity for intricate medical interventions. This project lays the foundation for a more equitable and data-informed healthcare system, paving the way for a future where personalized medicine empowers individuals and optimizes overall well-being. | |
| Role by  Member |  |
| 1- Data acquisition (all members )  2-Preprocessing of the data ( Ahmed and Abdelwahid)  3- Build and train the model (Madhumitha and Mustafa)  4- Evaluate the Performance (Taha and Ranim)  5- Deployment ( Extra if there is time ) (Ranim and Ahmed)  Note: all the members hold responsibility on each part, and they will help and give feedback to the main members assigned to it | |
| Schedule  Summary |  |
| **Important dates**   * 15 Jan 2024 : 1st Mentoring Live session * 18 Jan 2024 : Final Version of Capstone Action Plan Deadline * 22 Jan 2024 : 2nd Mentoring Live Session * 28 Jan 2024 : 3rd Mentoring Live Session * 31 Jan 2024 : Submission of Capstone Final Report, Presentation and Coding Reference * 1 Feb 2024 (6 PM -7:30) : CAPSTONE FINAL PITCH Live Session   **Detailed Schedule**   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Task** | **Start date** | **End date** | **Responsibility** | **Deliverable** | | Data acquisition | 10/1/2024 | 13/1/2024 | All members | Code  Data Report | | Preprocessing of the data | 14/1/2024 | 18/1/2024 | Ahmed Abdelwahid | Code  Data  Report | | Build and train the model | 19/1/2024 | 23/1/2024 | Madhumitha Mustafa | Code Report | | Evaluate the Performance | 24/1/2024 | 26/1/2024 | Taha  Ranim | Code Report | | Finalizing the Final report | 27/1/2024 | 28/1/2024 | All members | Report | | Working the PPT | 29/1/2024 | 30/1/2024 | All members | PPT | | Preparing for the demo | 31/1/2024 | 31/1/2024 | All members |  | | Submission of Capstone Final Report, Presentation and Coding Reference | 31/1/2024 | 31/1/2024 | All members | Report  Presentation Coding | | Demo | 1/2/2024 | 1/2/2024 | All members | PPT | | |
| Comment &  Assessment |  |
| <Comment and assessment **by the instructor.**> | |